

## Annexes

### Essais cliniques randomisés vs études d'observation

« Nous avons trois moyens principaux : l'observation de la nature, la réflexion et l'expérience. L'observation recueille les faits ; la réflexion les combine ; l'expérience vérifie le résultat de la combinaison. » – Denis Diderot. Pensées sur l'interprétation de la nature (§ XV, 1754) [65].

Les ECR, lorsqu'ils sont disponibles pour une question clinique, doivent être privilégiés, car ils minimisent les biais et fournissent des données plus robustes [66]. Récemment, Fanaroff et al. [67] ont discuté l'idée que l'observation clinique et le bon sens pourraient parfois être préférés aux ECR pour éclairer les décisions cliniques. Ils soulignent que les résultats des ECR bien conçus ont souvent contredit les pratiques basées sur le bon sens. Ils plaident pour que les ECR restent la référence pour établir des preuves solides, car ils fournissent des résultats plus fiables que les observations cliniques seules.

### Méta-analyse séquentielle (trial sequential analysis ou TSA)

La TSA est une méthode (assez récente) de mété-analyse cumulative, utilisée pour pondérer les erreurs de type I et II (liées à la répétition des tests statistiques), et pour estimer à partir de quel moment l'effet d'une intervention est suffisamment important pour ne pas être influencé par des études ultérieures.

Contrairement aux mété-analyses conventionnelles, la TSA permet donc d'ajuster sur les tests répétés et l'hétérogénéité, ce qui conduit à une réduction du risque de conclusion prématurée.

### Méta-analyse de second ordre

Selon Schmidt et al. [68], une mété-analyse de second ordre est une mété-analyse d'un certain nombre de mété-analyses de premier ordre, statistiquement indépendantes et méthodologiquement comparables, examinant la même relation dans différents contextes. La mété-analyse de premier ordre permet de réduire de façon significative la variance de l'erreur d'échantillonnage.

C'est donc une synthèse méthodologique qui agrège et analyse les résultats de plusieurs mété-analyses. L'unité d'analyse est la mété-analyse elle-même, et non les essais individuels.

Les objectifs théoriques sont d'évaluer la qualité méthodologique des mété-analyses existantes, d'en estimer la robustesse globale et d'identifier les discordances.

Les objectifs pratiques sont : résumer l'ensemble des mété-analyses portant sur un même thème ; comparer leurs conclusions ; estimer un effet global à par-

tir des effets agrégés ; évaluer la qualité et la concordance des revues.

Les analyses concernent le chevauchement (*overlap*) des études primaires entre les différentes mété-analyses, l'hétérogénéité entre mété-analyses (critères d'inclusion, définition des événements, méthodes statistiques), ainsi que la qualité des mété-analyses incluses. Les recommandations de bonnes pratiques et les étapes méthodologiques essentielles pour réaliser des mété-revues rigoureuses et systématiques sont disponibles dans l'article de Hennessy et al. [69].

### Détail des méthodes statistiques de la mété-analyse de la Société française d'hygiène hospitalière

#### Méta-analyse fréquentiste

Une approche fréquentiste a été mise en œuvre dans le cadre de cette mété-analyse [a]. Les effets observés dans les différentes études ont été combinés à l'aide d'un modèle à effets aléatoires (reposant sur l'hypothèse que les vrais effets diffèrent entre les études et fluctuent autour d'un effet moyen global, avec une variance inter-études notée  $\tau^2$ ).

La variance inter-études ( $\tau^2$ ) a été estimée à l'aide de la méthode de Sidik-Jonkman [b]. Cette méthode est robuste, en particulier en présence d'un nombre limité d'études et d'une hétérogénéité importante (hétérogénéité importante dans notre cas).

Les intervalles de confiance de l'effet combiné ont été calculés en utilisant la méthode de Hartung-Knapp (également appelée méthode de Hartung-Knapp-Sidik-Jonkman ou HKSJ). Celle-ci procure une inférence plus robuste et conservatrice que l'approche classique de Der-Simonian-Laird).

L'hétérogénéité entre les études a été évaluée via le test Q de Cochran. La statistique  $I^2$  de Higgins et Thompson [c] a été calculée pour quantifier la proportion de la variabilité totale attribuable aux différences réelles entre les études plutôt qu'à l'erreur d'échantillonnage. Lorsque cela était possible, un intervalle de confiance à 95% pour  $\tau^2$  a été calculé à l'aide de méthodes de profil basées sur la statistique Q.

Conformément aux recommandations d'IntHout et al. [d], nous présentons également l'intervalle de prédition à 95% (l'intervalle de confiance caractérise l'incertitude concernant l'effet moyen ; l'intervalle de prédition permet d'estimer l'intervalle dans lequel l'effet pourrait se situer dans une future étude).

Un diagramme en entonnoir (*funnel plot*) a été produit afin d'évaluer visuellement un éventuel biais de publication [e,f]. Les tests statistiques associés, tels que le test de régression d'Egger ou de Begg [g], ont été réalisés.

### Méta-analyse bayésienne

La réalisation d'une mét-a-analyse en présence de faibles taux d'événements, d'un petit nombre d'études, ou des deux simultanément, pose des difficultés méthodologiques. En particulier, l'estimation de l'hétérogénéité inter-études est délicate dans ce contexte, et une estimation imprécise de cette hétérogénéité peut conduire à des intervalles de confiance artificiellement trop étroits et à des taux d'erreur de type I inappropriés.

La SF2H a donc réalisé une mét-a-analyse bayésienne complémentaire, reposant sur une vraisemblance binomiale (en modélisant directement les nombres d'événements observés, selon une distribution binomiale), avec une fonction de lien logit. Cette approche présente l'avantage majeur d'éviter le recours à des corrections de continuité (telles que l'ajout de 0,5 dans les cellules vides) lorsque certaines études rapportent zéro événement dans un bras ou dans les deux bras de traitement (situation effectivement rencontrée dans notre mét-a-analyse).

Des analyses de sensibilité ont été conduites, comme recommandé par Rosenberger et al. [h], afin d'évaluer l'influence du choix de différents hyperparamètres ou de différents types de distributions a priori sur les conclusions de la mét-a-analyse. Cette démarche permet d'évaluer la robustesse des résultats obtenus. Nous avons utilisé des distributions a priori faiblement informatives pour le paramètre d'hétérogé-

néité  $\tau$ , conformément aux propositions de Günhan et al. et de Röver et al. [i,j]. Cela permet d'incorporer une information minimale, tout en laissant les données déterminer principalement l'estimation de l'hétérogénéité. Nous avons initialement utilisé les a priori suivants: une distribution t de Student (0 ; 0,66 ; 3) pour l'effet et une distribution inverse gamma (2,08 ; 0,86), pour l'hétérogénéité.

La convergence des chaînes de Markov Monte-Carlo (MCMC) utilisées pour l'estimation bayésienne a été évaluée à l'aide du diagnostic de Gelman et Rubin [k], également connu sous le nom de statistique R-hat. Les intervalles de confiance (approche fréquentiste) et les intervalles de crédibilité (approche bayésienne) ont été calculés selon un niveau conventionnel de 95%, permettant ainsi une comparaison directe entre les deux approches méthodologiques.

Les analyses fréquentistes ont été réalisées à l'aide du logiciel Stata® (version 19.5) ainsi que de plusieurs packages du langage R® (notamment les packages meta et metafor).

Les analyses bayésiennes ont été conduites en utilisant plusieurs packages du langage R® (version 4.5.5) (par exemple RoBMA, brms, rstanarm).

Stata® (StataCorp, College Station, TX, États-Unis): <https://www.stata.com/>

R® (The R Foundation for Statistical Computing, Vienne, Autriche): <https://cran.r-project.org/> ■

### Références méthodologiques

- a- Seide SE, Jensen K, Kieser M. A comparison of Bayesian and frequentist methods in random-effects network meta-analysis of binary data. *Res Synth Methods*. 2020;11(3):363-378.
- b- Sidik K, Jonkman JN. Simple heterogeneity variance estimation for meta-analysis. *J R Stat Soc C: Appl Stat*. 2005;54:367-384. DOI: 10.1111/j.1467-9876.2005.00489.x.
- c- Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21(11):1539-1558.
- d- IntHout J, Ioannidis JPA, Rovers MM, et al. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*. 2016;6(7):e010247.
- e- Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629-634.
- f- Chaimani A, Higgins JPT, Mavridis D, et al. Graphical tools for network meta-analysis in STATA. *PLoS One*. 2013;8(10):e76654.
- g- Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics*. 1994;50(4):1088-1101.
- h- Rosenberger KJ, Xing A, Murad MH, et al. Prior choices of between-study heterogeneity in contemporary Bayesian network meta-analyses: an empirical study. *J Gen Intern Med*. 2021;36(4):1049-1057.
- i- Günhan BK, Röver C, Friede T. Random-effects meta-analysis of few studies involving rare events. *Res Synth Methods*. 2020;11(1):74-90.
- j- Röver C, Bender R, Dias S, et al. On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. *Res Synth Methods*. 2021;12(4):448-474.
- k- Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci*. 1992;7(4):457-472.